Microbiology Research International
Vol. 10(3), pp. 25-31, November 2022
DOI: 10.30918/MRI.103.22.022
ISSN: 2354-2128
Review

Illumina sequencing technology review

Teferi Benti Moti

Department of Veterinary Micobiology, Animal Health Institute, Sebeta, Ethiopia.

Accepted 25 October, 2022

ABSTRACT

As a result of Sanger's original DNA sequencing discovery, rapidly evolving DNA sequencing systems demonstrate how biology and technology may harmoniously coexist. Prior to 2005, Sanger sequencing technology served as the foundation for all sequencing. Next-generation sequencing (NGS) is a breakthrough technology that promises to advance our knowledge of how nucleic acid functions significantly. The DNA sequencing process is to determine the proper arrangement of nucleotide bases in a DNA macromolecule using biological methods. The development of Illumina-Next Generation Sequencing Technologies in 2005 revolutionized how scientists study and comprehend biological phenomena. In the last ten years, new sequencing devices have advanced dramatically. Science is developing very quickly, and the most popular subject in the field of human and animal genomics research right now is Illumina's next-generation sequencing technology. In this paper, we have underlined a number of fundamental principles, benefits, and comprehensions of next-generation sequencing technologies. The developments in the sequencing platforms will be given, together with a brief explanation of first, second, and third generation sequencing technologies.

Keywords: DNA, Illumina, next-generation sequencing, Sanger sequencing.

Email: teferibenti58@gmail.com. Tel: +251911023089.

INTRODUCTION

The discovery of the double helix structure composed of four deoxyribonucleic acid (DNA) base A, T, C, and G by Watson and Crick (1953) led to the decoding of genomic sequences and the knowledge of the DNA composition of organisms. DNA sequencing is a discovery that uses the DNA structure to understand and decrypt the code for all biological life on earth as well as to understand and treat genetic diseases. The appearance of sequencing technologies has played an important role in the analysis of genomic sequences of organisms (Le Tourneau and Kamal, 2015).

A DNA sequencer produces files containing DNA sequences (Shendure and Ji, 2008). DNA sequencing technologies have existed since the early 1970s, but initially, their cost, complexity, and requirement for toxic or radioactive reagents limited their use to research settings (Lander, 2011). The chain-termination methods pioneered by Sanger and colleagues (Coulson et al., 1998) were more practical and formed the basis for the first generation of automated DNA sequences. Public health applications were first introduced in the 1990s, such as the multilocus sequence typing scheme for Neisseria meningitides developed by (Maiden et al., 1998). The first complete genome of a free-living microorganism, Haemophilus influenza, published in 1995, was sequenced using the Sanger method (Fleischmann et al.,1995).

The Human Genome Project (HGP), which was started in 1990 and aims to sequence and understand the 3.2 billion nucleotide base pairs that make up the human genome for possible medicinal advantages, has increased the need for high-speed sequencing technologies. This $3.8 billion international partnership was founded initially on Gilbert WA's Sanger sequencing work (1977). However, various new techniques have been directly developed from the improvements made to Sanger's approach because whole genome sequencing (WGS) utilizing this technology is extremely expensive, time-consuming, and low in output and accuracy (Tripp and Grueber,2011). Second-generation sequencing

platforms require amplified sequencing libraries, making it time-consuming and costly, while third-generation single-molecule sequencing can be undertaken without the requirement of amplified DNA clones. When new next-generation sequencing (NGS) techniques using massively parallel processing brought the cost down to a tiny fraction of the price of Sanger sequencing and lowered the sequencing time, a significant change occurred in the early 2000s.

Whole genome sequencing (WGS) for bacterial pathogens has been migrating from research laboratories into public health practice since the 2010s. Whole genome sequencing (WGS) analyses were conducted on cultured isolates. Whole genome sequencing has the potential to rapidly provide a large amount of information from isolates, including species, strain type, antibiotic resistance, virulence, and other information for outbreak and case management. While the value of whole genome sequencing (WGS) for outbreak detection and investigation is clear in many situations, at current cost levels, the usefulness of this approach is less clear for the diagnosis and treatment of individual patients, especially considering the emergence of direct-from-specimen multi-analyte test panels. These tests have the ability to identify common pathogens in patient specimens in a highly useful time frame (Vezzi, 2012). Therefore, the objective of this review is to highlight the historical background and principles of Illumina Sequencing Technology (next-generation sequencing).

REVIEW

The first generation of sequencing

The primary next-generation sequencing technologies were the Sanger and Maxam-Gilbert technologies (Thudi et al., 2012). Allan Maxam and Walter Gilbert developed Maxam-Gilbert sequencing in 1977–1980 and it's known as the chemical method of DNA sequencing, while Sanger sequencing was developed (Sanger et al, 1977).

Maxam-Gilbert sequencing

Maxam-Gilbert sequencing is one of the two conventional DNA sequencing methods and is the first generation of sequencing. It relies on the cleaving of nucleotides by chemicals and is best with small nucleotide polymers. Chemical treatment generates breaks at a small proportion of one or two of the four nucleotide bases in each of the four reactions (C, T+C, G, A+G). This reaction results in a series of marked fragments that can be separated according to their size by electrophoresis (El-Metwally et al., 2014). The sequencing here is performed without DNA cloning. However, the events and improvement of the Sanger sequencing method favored the latter over the Maxam-Gilbert sequencing method. This method is more sensitive and specific. However, it is also considered hazardous because it uses toxic and radioactive chemicals.

Sanger sequencing

Sanger sequencing is the alternate conventional DNA sequencing system, which is extensively used. Generally, it uses labeled ddNTPs to terminate the chain growth during DNA replication at each of the four nucleotides. Eventually, the separation of the terminated amplicons on a gel allows the determination of the DNA sequence. Sanger sequencing is known as the chain termination system or the sequencing by synthesis method. One strand of the double-stranded DNA is used as a template to be sequenced. Sanger sequencing is made using chemically modified nucleotides called dideoxy-nucleotides (dNTPs).

Dideoxynucleotides are used to extend nucleotides using dNTPs. They stop further extension after being integrated into the DNA strand, marking the end of the elongation process. Then, we have DNA fragments of various sizes that have been terminated by a dNTP. The fragments are separated based on their size using a gel slab, and the resulting bands can be seen using an imaging device (X-ray or UV light) (El-Metwally et al., 2014; Masoudi-Nejad et al., 2013). Importantly, Sanger sequencing is a considerably streamlined DNA sequencing technique. As a result, DNA sequencing received a boost with the invention of the technology, enabling a faster accumulation of sequence data for various genes and organisms. The Sanger sequencing method's sensitivity is still quite low, and the sequencing process is still very costly and time-consuming. Chemicals used during the procedure are less hazardous.

Second generation of sequencing

The Sanger method ("first generation" technology) was the primary sequencing technology between 1975 and 2005 and produces relatively long (500 to 1000 bp) reads. However, the cost and time were major stumbling blocks. The emergence of a new generation of sequencers to break the limitations of the first generation has been marked. The basic principle on which the new generation of sequencing (NGS) works is similar to traditional Sanger sequencing methods involving capillary electrophoresis. Different Illumina-NGS platforms adopt their own specific protocols and sequencing methods. The introduction of pyrosequencing technology by Roche

454 Life Sciences in 2005 began the "next-generation sequencing" revolution (Martin and Goldstein, 2014). This high throughput technology allowed the generation and detection of thousands to millions of short sequencing reads in a single machine run without the need for cloning. Next-generation sequencing technologies have emerged that generate both short (50 to 400 bp) and long (1 to 100 kb) reads (Chen et al., 2015). The short-read technologies currently in use are collectively known as massively parallel sequencing and are often also referred to as second-generation sequencing (Long et al., 2016). They produce billions of nucleotide sequences during each run, where each genome is sequenced multiple times in small random pieces to generate very large data sets and platforms have different biochemistry and arrays. The basic characteristics of second-generation sequencing technology are:

• The simultaneous generation of many millions of short reads,
• The acceleration of the sequencing process in comparison to the first generation,
• The low cost of sequencing, and
• The sequencing output is directly detected without the need for electrophoresis.

The Illumina New Generation Sequence (NGS) procedures consist of four fundamental steps: The beginning material is thought to be double-stranded DNA. However, the source from which this information is taken may be genomic DNA, immuno-precipitated DNA, reverse-transcribed RNA, or cDNA (Rizzo and Buck, 2012).

Library preparation: the sequencing library is prepared by random fragmentation of the DNA or cDNA sample, followed by 5′and 3′adapter ligation. Alternatively, “tagmentation” combines the fragmentation and ligation reactions into a single step that greatly increases the efficiency of the library preparation process. Adapter-ligated fragments are then PCR amplified and gel purified.

Various standard library preparation kits offer protocols for whole-genome sequencing (WGS), RNA sequencing (RNA-Seq), targeted sequencing (such as exome sequencing or 16S sequencing), custom-selected regions, and protein-binding regions. Although the number of new generation sequencing methods is constantly growing, in 2008, Illumina introduced an upgrade, the Genome Analyzer II, which triples output compared to the previous Genome Analyzer instrument. Moreover, sequence library preparation involves some common steps of fragmentation of DNA templates into smaller pieces, with specific size selection depending upon the requisite platform. Additionally, adapter ligation is also involved in this process, which adds platform-specific, synthetic DNA at the end of the DNA fragments present in this library to facilitate the sequencing reactions.

Cluster generation: For cluster generation, the library is loaded into a flow cell where fragments are captured on a lawn of surface-bound oligos complementary to the library adapters. Each fragment is then amplified into distinct, clonal clusters, through bridge amplification when cluster generation is complete; the templates are ready for sequencing. This step involves either the attachment of a DNA fragment to a microbead or the same to a glass slide when some PCR techniques are followed. Library amplification eventually leads to the sequencing reaction and imaging process. It involves the assessment of some important and vital genes or regulatory elements in the given genome.

Sequencing by synthesis (SBS) technology from Illumina employs a proprietary reversible terminator-based method for detecting single bases as they are incorporated into DNA template strands. Natural competition minimizes incorporation bias and greatly reduces raw error rates compared to other technologies because all four reversible terminator-bound dNTPs are present during each sequencing cycle. The result is highly accurate base-by-base sequencing that virtually eliminates sequence context-specific errors, even within repetitive sequence regions and homopolymers (Ross et al., 2013; Bentley et al., 2008).

Data analysis

During data analysis and alignment, the newly identified sequence reads are aligned to a reference genome. Following alignment, many variations of study are possible, like single nucleotide polymorphism (SNP) or insertion-deletion identification, read counting for RNA methods, phylogenetic or metagenomic analysis, and more. A close animation of Sequencing by synthesis (SBS) chemistry is obtainable at (www.illumina.com/SBS) Illumina, Inc. (2015). The workflow to investigate large-scale new generation sequencing (NGS) should include the following: data quality assessment, comprehensive analysis, interpretation of results, and presentation of knowledge in a very meaningful format. Several free and commercial software tools are available for the analysis and visualization of new-generation sequencing (NGS) data.

Short-read sequencing approaches are divided into two broad categories: sequencing by ligation (SBL) and sequencing by synthesis (SBS), according to Baker et al. (2016). Second-generation sequencing is classified into three major sequencing platforms: A detailed animation of Sequencing by synthesis (SBS) chemistry is available at www.illumina.com/SBS Illumina, Inc. (2015).

• Roche/454, introduced in 2005.
• In 2006, Illumina Sequencing and
• The ABI/SOLiD in 2007.

Roche/454 sequencing

DNA samples are randomly fragmented and each fragment is attached to a bead whose surface carries primers that have oligonucleotides complementary to the DNA fragments, so each bead is associated with single fragments. Each bead is isolated and amplified using PCR emulsion, which produces about one million copies of each DNA fragment on the surface of the bead. The beads are then transferred to a plate containing many wells called a Pico titer plate (PTP) and the pyrosequencing technique is applied, which consists of activating a series of downstream reactions, producing light at each incorporation of a nucleotide. The sequence of the DNA fragment is deduced by detecting the light emission after each nucleotide incorporation (Mardis, 2008).

In 2008, Roche released the upgraded 454 GS FLX Titanium system with an average read length of 700 bp, 99.997% accuracy, and 0.7 Gb of data per run output within 24 hrs, which is easier to map to a reference genome. The major drawback of Roche 454 in most cases was the high error rates during sequencing insertions and deletions due to the presence of homopolymer regions (Huse et al*.,* 2007). Indeed, the identification of the size of homopolymers should be determined by the intensity of the light emitted by pyrosequencing. Signals with too high or too low-intensity lead to under or over-estimation of the number of nucleotides, which causes errors in nucleotide identification.

ABI/SOLiD sequencing

Sequencing by oligonucleotide ligation and detection systems was first released by Applied Biosystems Instruments (ABI) in 2008. Based on 2-nucleotide sequencing by ligation (SBL), Le Tourneau and Kamal (2015) describe the sequential annealing and subsequent ligation of probes to the template. The Applied Biosystems Instruments (ABI/SOLiD process consists of multiple sequencing rounds. It starts by attaching adapters to the DNA fragments, fixed on beads and cloned by PCR emulsion. These beads are then placed on a glass slide, and the 8-mer with a fluorescent label at the end is sequentially ligated to DNA fragments, and the color emitted by the label is recorded. The output format is color space, which is the encoded form of the nucleotide where four fluorescent colors are used to represent 16 possible combinations of two bases (Le Tourneau and Kamal, 2015).

The sequencer repeats this ligation cycle, and in each cycle the complementary strand is removed and a new sequencing cycle starts at position n-1 of the template. The cycle is repeated until each base is sequenced twice. ABI/SOLiD launched the first sequencer that produced short reads (Long et al., 2016), increased the length of reads to 35 bp with an output of 3 Gb/run and continued to improve their sequencing, which increased the length of reads to 75 bp with an output of up to 30 Gb/run. The strength of the ABI/SOLiD platform is high accuracy because each base is read twice, while the drawback is a relatively short read and long run times. The errors in sequencing in this technology are due to noise during the ligation cycle, which causes errors in the identification of bases. The main type of error is substitution (Kchouk et al., 2017).

Illumina/Solexa / HiSeq and MiSeq *sequencing*

The Solexa sequencing platform was commercialised in 2006, with Illumina acquiring Solexa in early 2007. The principle is based on sequencing-by-synthesis chemistry, with novel reversible terminator nucleotides for the four bases each labeled with a different fluorescent dye, and a special DNA polymerase enzyme able to incorporate them. During the first step, the DNA samples are randomly fragmented into sequences and adapters are ligated to both ends of each sequence. These adapters are fixed themselves to the respective complementary adapters; the latter are hooked on a slide with many variants of complementary adapters placed on a solid plate. During the second step, each attached sequence to the solid plate is amplified by "PCR bridge amplification" which creates several identical copies of each sequence; a set of sequences made from the same original sequence is called a cluster. Each cluster contains approximately one million copies of the same original sequence (Shendure and Ji, 2008).

The last step is to determine each nucleotide in the sequences. Illumina uses the sequencing by synthesis approach that employs reversible terminators (Bentley et al*.,* 2008), in which the four modified nucleotides, sequencing primers, and DNA polymerases are added as a mix, and the primers are hybridized to the sequences. Then, polymerases are used to extend the primers using the modified nucleotides. Each type of nucleotide is labeled with a fluorescent specific in order for each type to be unique. The nucleotides have an inactive 3'-hydroxyl group, which ensures that only one nucleotide is incorporated. Clusters are excited by a laser to emit a light signal specific to each nucleotide, which will be detected by a coupled-charge device camera, and computer programs will translate these signals into a nucleotide sequence.

The process continues with the elimination of the terminator with the fluorescent label and the start of a

new cycle with a new incorporation (Chen et al., 2015). The first sequencers, Illumina/Solexa/GA, were able to produce very short reads of 35 bp, and they had an advantage in that they could produce paired-end (PE) short reads, in which the sequence at both ends of each DNA cluster is recorded. The output data of the last Illumina sequencers is currently higher than 600 Gpb and the lengths of short reads are about 125 bp (Kulski, 2016). Illumina provides at least eight industrial-level sequencing machines (NextSeq 500, HiSeq series 2500, 3000 and 4000, and HiSeq X series five and ten) having mid-to-high output (120 to 1500 Gb) (Liu et al., 2012). One of the main drawbacks of the Illumina/Solexa platform is the high requirement for sample loading control because overloading can result in overlapping clusters and poor sequencing quality. The overall error rate of this sequencing technology is about 1%. Substitutions of nucleotides are the most common type of error in this technology (Dohm et al., 2008).

The automated short-read sequencing platforms differ substantially in terms of their engineering, sequencing chemistry, output (length of reads, number of sequences), accuracy, and price (Heather and Chain, 2015). Masoudi-Nejad et al. (2013) explains that the Illumina platform, which currently occupies an infinite part of the next-generation sequencing (NGS) market, relies on sequencing by synthesis of the complementary strand and fluorescence-based detection of reversibly-blocked terminator nucleotides (Masoudi-Nejad et al., 2013). The platform includes multiple instruments with varying throughput and read length. All of the enzymatic processes and imaging steps of the Illumina technology happen during a very long flow cell. Looking at the current Illumina platform, it should be partitioned into 1 (miSeq), 2 (HiSeq2500), or 8 (HiSeq2000, HiSeq2500) separate lanes. The Illumina platform uses bridge amplification for polony generation and a sequencing by synthesis (SBS) approach (Heo, 2015).

In 2008, an upgrade, the Genome Analyzer II, was announced that triples output compared to the previous Illumina Genome Analyzer device. A paired-end module for the sequencer was introduced, and with new optics and camera components that allow the system to image DNAclusters more efficiently over larger areas, the new instrument triples the output per paired-end run from 1 to 3 Gb. The system generates a minimum of 1.5 Gb of single-read data per run, and a minimum of three Gb of data in an exceedingly paired-end run, recording data from quite 50 million reads per flow cell. The run time for a 36-cycle run was decreased to 2 days for a single-read run and 4 days for a paired-end run. Typical descriptions of the Genome Analyzer system can be found at http://www.solexa.com/ and in Schuster (2008).

The creators of 454 sequencing launched the Ion Torrent technique with two key modifications (Rothberg et al., 2011). The pH shift in the surrounding solution is electronically detected proportionate to the quantity of inserted nucleotides, as opposed to a weak signal being recognized by an optical device. The sequencing reactions are carried out inside a microchip combined with flow cells, while electronic sensors are mounted on the underside of the flow cells. The Ion Personal Genome Machine (PGM), a bench-top sequencer with 11.1 million sensors, and the high-throughput Proton sequencer, with over 165 million sensors, are the commercial sequencers that make use of Ion Torrent technology (Shiina et al., 2018). It has been developed to detect individual protons using an ion-sensitive field-effect transistor sensor. The chip is flushed with unlabeled dNTPs in the presence of DNA polymerase when it is put inside the flow cell. The integration of the nucleotide into the DNA chain releases a proton, which causes a pH change that can be felt. The main drawbacks were problems reading homopolymer repetitions, whereas the various benefits include considerably longer read lengths, lower costs, and shorter workdays (Diekstra et al., 2015).

Third-generation sequencing

Third regeneration sequencing immediately offers three key advantages over the second generation: it generates long reads, it is fast, and it is simple. The availability of long reads will have a major impact on evolutionary studies involving assembly or barcoding approaches. Assembling genomes solely supported by short reads with no available references remains a challenge (Koren and Phillippy, 2015). The third generation of sequencing techniques recently became available either commercially or, at a minimum, for selected beta-testers. Pacific Biosystems (PacBio) uses Single Molecule Real-Time (SMRT) sequencing (Eid et al., 2009), whereas Oxford Nanopore Techniques (ONT) developed a tool for nanopore sequencing (Branton et al., 2008), Helicos True Single Molecule Sequencing, and in contrast to second-generation techniques (454, Illumina, Ion Torrent), these methods do not include an amplification step during sequencing, or shorter library preparation times while decreasing the probabilities of error and enabling single molecule sequencing. This enables billions of unique fragments to be independently sequenced at the same time. Moreover, the expected read lengths are much longer than those of second-generation techniques, with average read lengths exceeding 6 to 8 kbps and maximal read lengths exceeding 30 to 150 kbps.

Helicos sequencing

The Helicos sequencing provider Seqll sequences the genomic DNA and RNA by means of the Helicos sequencing system and HeliScope single-molecule

sequencers (Shendure and Ji, 2008). This method is an amalgamation of sequencing by hybridization and sequencing by synthesis employing a DNA polymerase. Sheared DNA is tailed with polyA and hybridized to a disposable glass flow cell surface surrounded by oligo-dT, allowing the parallel sequencing by synthesis of billions of molecules. The strategy of adding fluorescent nucleotides with a terminating nucleotide will pause the tactic until each nucleotide of the DNA sequence has been captured. The Helicos sequencing process repeats itself until all the fragments are sequenced completely (Eid et al., 2009). It avoids size bias or GC content bias since there is no requirement for PCR amplification or ligation as compared with other methods and the lengths of sequencing reads range from 25 to 60 bases (Hart et al., 2010).

Single-molecule real-time (SMRT) sequencing

Pacific Biosciences has developed the sole Molecule Real Time DNA Sequencing (SMRT) and markets the PacBio RS II sequencer (Schadt et al., 2010). The SMRT sequencer uses Zero mode waveguides (ZMWs) which contain 1.5 lac ultra-microwells where each molecule of DNA polymerase is immobilized at the underside of the well using the biotin-streptavidin system. Once the coupling of the single-stranded template DNA with the immobilized polymerase takes place, each nucleotide incorporation is detected by the addition of the fluorescently labeled dNTP analogs. Zero mode waveguides (ZMWs) are continuously monitored using CCD cameras, and a series of pulses are converted into single molecular traces equivalent to the template sequence. This platform of sequencing allows for a faster genome assembly than comparative technologies since all four nucleotides are added simultaneously and measured in real-time. 99.3% accuracy has been reported with a read length of 900 bp (Metzker, 2010).

Nanopore sequencing (MinION and PromethION)

The most recent single-molecule sequencing techniques were introduced in 2012 by Oxford Nanopore Technologies. The PromethION system might be a compact bench top, whilst the MinION Mkl sequencing machine could be a portable device the size of a USB drive that attaches to a PC or laptop for DNA and RNA sequencing. The whole concept behind nanopores is that each individual nucleotide will change the ionic current as it passes through the pore, producing time-specific signals that are then assessed on a real-time basis (Bayley, 2015).

The type of molecule moving through the pore will control how much ion current flows (Stoddart et al., 2009). The minimal sample preparation, lack of PCR amplification or ligation steps, and longer read durations are the key benefits of this method (kbp range). The main challenge appears to be improving the speed of DNA translocation through the nanopores. As a result of the high error rates being reduced, this may further ensure accurate measurements of the current changes (Bayley, 2015).

CONCLUSION

The first method of sequencing came about half a century ago, and since then, sequencing technologies have continued to evolve, particularly after the appearance of the first sequencers from New Generation Sequencing Technology, which appeared in 2005. Technical developments make it feasible to improve technical nucleic acid sequencing abilities from the preliminary Sanger technique to actual next-generation sequencing. Currently, Illumina's next-generation sequence (NGS) technology is one of the most potent and successful methods for quick sequencing of DNA and RNA. In comparison to the first generation of sequencing technology, the second generation has revolutionized DNA analysis and is the most often utilized. However, the PCR amplification steps are time-consuming and expensive in terms of sequencing cost, and the relatively short reads make genome assembly more problematic. The third-generation approaches have no amplification phase during the sequencing process, which reduces the possibility of mistakes and enables single-molecule sequencing. The third-generation sequencing methods allow for the low-cost assembling of billions of reads while maintaining high throughput. These technologies are currently the foundation for a number of research fields that facilitate the investigation and evaluation of biological sequences. Despite this, many challenges remain with Illumina-NGS technologies in terms of huge data acquisition and storage, data analysis and interpretation. New next-generation sequencing platforms will emerge in the upcoming years, producing a far higher volume of data (measured in terabytes) that requires the assistance of cutting-edge techniques and software programs capable of processing this enormous volume of data.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

REFERENCES

Baker EAG, Goodwin S, McCombie WR, Ramos OM, **2016**. Silico: a simulator of long read sequencing in PacBio and Oxford Nanopore. BioRxiv, p.076901.

Bayley H, **2015**. Nanopore sequencing: From imagination to reality. Clin Chem 61: 25-31.

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, **2008**. Accurate whole human genome sequencing using reversible terminator chemistry. Nature, 456: 53–59.

Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, Di Ventra M, Garaj S, Hibbs A, Huang X, Jovanovich SB, Krstic PS, Lindsay S, Ling XS, Mastrangelo CH, Meller A, Oliver JS, Pershin YV, Ramsey JM, Riehn R, Soni GV, Tabard-Cossa V, Wanunu M, Wiggin M, Schloss JA, **2008**. The potential and challenges of nanopore sequencing. Nat Biotechnol, 26(10): 1146-53. doi: 10.1038/nbt.1495.

Chen R, Im H, Snyder M, **2015**. Whole-exome enrichment with the agilent sureselect human all exon platform. Cold Spring Harbor Protocols, 2015(7): pdb-prot083659

Coulson TN, Pemberton JM, Albon SD, Beaumont M, Marshall TC, Guinness FE, Clutton-Brock TH, **1998**. Microsatellites reveal heterosis in red deer. Proceedings of the Royal Society of London. Series B: Biological Sciences, 265(1395): 489-495.

Diekstra A, Bosgoed E, Rikken A, van Lier B, Kamsteeg EJ, Tychon M, Derks RC, van Soest RA, Mensenkamp AR, Scheffer H, Neveling K, Nelen MR, **2015**. Translating sanger-based routine DNA diagnostics into generic massive parallel ion semiconductor sequencing. Clin Chem, 61(1): 154-62. doi: 10.1373/clinchem.2014.225250.

Dohm JC, Lottaz C, Borodina T, Himmelbauer H, **2008**. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Res, 36: e105.

Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S, **2009**. Real-time DNA sequencing from single polymerase molecules. Science, 323(5910): 133-138. doi: 10.1126/science. 1162986.

El-Metwally S, Ouda OM, Helmy M, **2014**. Next generation sequencing technologies and challenges in sequence assembly. Springer.

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, **1995**. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science, 269(5223): 496–512.

Hart C, Lipson D, Ozsolak F, Raz T, Steinmann K, Thompson J, Milos PM, **2010**. Single-molecule sequencing: sequence methods to enable accurate quantitation. Methods Enzymol, 472: 407-30. doi: 10.1016/S0076-6879(10)72002-4.

Heather JM, **Chain** B, **2015**. The sequence of sequencers: The history of sequencing DNA. Genomics, 107: 1-8.

Heo Y, **2015**. Improving quality of high-throughput sequencing reads. PhD Dissertation, University of Illinois at Urbana-Champaign.

Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM, **2007**. Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol, 8: R143.

Kchouk M, Gibrat JF, Elloumi M, **2017**. Generations of Sequencing Technologies: From First to Next Generation. Biol Med (Aligarh), 9: 395. doi:10.4172/0974-8369.1000395.

Koren S, **Phillippy** AM, **2015**. One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. Curr Opin Microbiol, 23: 110-120.

Kulski J ed., **2016**. Next generation sequencing: Advances, applications and challenges. ISBN 978-953-51-2240-1.

Lander ES, **2011**. Initial impact of the sequencing of the human genome. Nature, 470(7333): 187–197.

Le Tourneau C, **Kamal** M, **2015**. Pan-cancer integrative molecular portrait towards a new paradigm in precision medicine. https://hal.archives-ouvertes.fr/hal-02560964/.

Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M, **2012**. Comparison of next-generation sequencing systems. J Biomed Biotechnol, 2012:251364. doi: 10.1155/2012/251364.

Long RG, Bürki A, Zysset P, Eglin D, Grijpma DW, Blanquer SB, Hecht AC, Iatridis JC, **2016**. Mechanical restoration and failure analyses of a hydrogel and scaffold composite strategy for annulus fibrosus repair. Acta Biomater, 30: 116-125.

Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, **1998**. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A, 95(6): 3140–5.

Mardis ER, **2008**. Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet, 9: 387-402.

Martin AC, **Goldstein** B, **2014**. Apical constriction: themes and variations on a cellular mechanism driving morphogenesis. Development, 141(10): 1987-1998.

Masoudi-Nejad A, Narimani Z, Hosseinkhan N, **2013**. Emergence of Next-Generation Sequencing. In Next Generation Sequencing and Sequence Assembly (pp. 11-39). Springer, New York, NY.

Metzker ML, **2010**. Sequencing technologies - The next generation. Nature Rev Genet, 11: 31-46.

Rizzo JM, **Buck** MJ, **2012**. Key Principles and Clinical Applications of "Next-Generation" DNA Sequencing. Cancer Prev Res (Phila), 5(7): 887-900. doi: 10.1158/1940-6207.CAPR-11-0432.

Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M, Dimalanta E, Dressman D, Kasinskas R, Sokolsky T, Fidanza JA, Namsaraev E, McKernan KJ, Williams A, Roth GT, Bustillo J, **2011**. An integrated semiconductor device enabling non-optical genome sequencing. Nature, 475(7356):348-52. doi: 10.1038/nature10242.

Sanger F, Nicklen S, Coulson AR, **1977**. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A, 74(12): 5463–7.

Schadt EE, Turner S, Kasarskis A, **2010**. A window into third-generation sequencing. Hum Mol Genet, 19(R2): R227-40. doi: 10.1093/hmg/ddq416.

Schuster SC, **2008**. Next-generation sequencing transforms today's biology. Nat Methods, 5(1): 16-8. doi: 10.1038/nmeth1156.

Shendure J, **Ji** H, **2008**. Next-generation DNA sequencing. Nature Biotechnol, 26: 1135-1145.

Shiina T, Suzuki S, Kulski JK, Inoko H, **2018**. Super high resolution for single molecule-sequence-based typing of classical HLA loci using Ion Torrent PGM. In *HLA Typing* (pp. 115-133). Humana Press, New York, NY.

Stoddart D, Heron AJ, Mikhailova E, Maglia G, Bayley H, **2009**. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. Proc Natl Acad Sci U S A, 106(19): 7702-7. doi: 10.1073/pnas.0901054106.

Thudi M, Li Y, Jackson SA, May GD, Varshney RK, **2012**. Current state-of-art of sequencing technologies for plant genomics research. Brief Funct Genomics, 11: 3-11.

Tripp S, **Grueber** M, **2011**. Economic impact of the human genome project. Battelle Memorial Institute; www.battelle.org.

Vezzi F, **2012**. Next generation sequencing revolution challenges: Search, assemble, and validate genomes. Ph.D, Universita degli Studi di Udine.

Watson JD, **Crick** FH, 1953. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. Nature 171: 737-738.

Citation: Moti TB, 2022. Illumina sequencing technology review. Microbiol Res Int, 10(3): 25-31.